



Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma

Citation

Sausen, M., R. J. Leary, S. Jones, J. Wu, C. P. Reynolds, X. Liu, A. Blackford, et al. 2012. "Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma." *Nature genetics* 45 (1): 12-17. doi:10.1038/ng.2493. <http://dx.doi.org/10.1038/ng.2493>.

Published Version

doi:10.1038/ng.2493

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:11717533>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Published in final edited form as:

Nat Genet. 2013 January ; 45(1): 12–17. doi:10.1038/ng.2493.

Integrated genomic analyses identify *ARID1A* and *ARID1B* alterations in the childhood cancer neuroblastoma

Mark Sausen^{1,11}, Rebecca J. Leary^{1,11}, Siân Jones^{1,9}, Jian Wu^{1,10}, C. Patrick Reynolds², Xueyuan Liu³, Amanda Blackford⁴, Giovanni Parmigiani^{5,6}, Luis A. Diaz Jr.¹, Nickolas Papadopoulos¹, Bert Vogelstein^{1,7}, Kenneth W. Kinzler¹, Victor E. Velculescu¹, and Michael D. Hogarty^{3,8}

¹Ludwig Center for Cancer Genetics and Therapeutics, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21287, USA

²Cancer Center, Texas Tech University Health Sciences Center, Lubbock, TX 79430, USA

³Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

⁴The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA

⁵Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA

⁶Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

⁷Howard Hughes Medical Institutions, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21287, USA

⁸Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

Correspondence should be addressed to V.E.V. (velculescu@jhmi.edu) or M.D.H. (hogartym@email.chop.edu).

⁹Current Address: Personal Genome Diagnostics, Inc., Science and Technology Park at Johns Hopkins, Baltimore, MD 21205, USA.

¹⁰Current Address: State Key Laboratory of Cancer Biology, Cell Engineering Research Center, The Fourth Military Medical University, Xi'an, P. R. China.

¹¹These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

M.S. and R.J.L. are joint first authors. C.P.R. established cell lines and C.P.R. and X.L. purified DNA samples from which M.S. prepared next generation DNA sequencing libraries. J.W. performed *MYCN* capture of genomic DNA libraries for massively parallel sequencing. M.S. and R.J.L. analyzed sequencing data for structural alterations. M.S., S.J., N.P., B.V., K.W.K., and V.E.V. sequenced next-generation DNA libraries and performed mutational analyses. A.B., G.P., and L.A.D. performed statistical analyses of clinical and sequencing data. M.S., R.J.L., B.V., K.W.K., V.E.V., and M.D.H. conceived the research and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

L.A.D., N.P., B.V., K.W.K., and V.E.V. are founders of Inostics and Personal Genome Diagnostics and are members of their Scientific Advisory Boards. L.A.D., N.P., B.V., K.W.K., and V.E.V. own Inostics and Personal Genome Diagnostics stock, which is subject to certain restrictions under university policy. The terms of these arrangements are managed by Johns Hopkins University in accordance with its conflict-of-interest policies.

URLs

1000 Genomes, <http://browser.1000genomes.org/index.html> (2010 release); dbSNP130 and dbSNP135, <http://www.ncbi.nlm.nih.gov/projects/SNP/>

ACCESSION NUMBERS

Sequence data have been deposited at the European Genome-Phenome Archive (EGA, <http://www.ebi.ac.uk/ega/>) which is hosted at the EBI, under accession number EGAS00001000369. Expression data have been previously deposited at the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE3960.

Neuroblastomas are tumors of peripheral sympathetic neurons and are the most common solid tumor in children. To determine the genetic basis for neuroblastoma we performed whole-genome sequencing (6 cases), exome sequencing (16 cases), genome-wide rearrangement analyses (32 cases), and targeted analyses of specific genomic loci (40 cases) using massively parallel sequencing. On average each tumor had 19 somatic alterations in coding genes (range, 3–70). Among genes not previously known to be involved in neuroblastoma, chromosomal deletions and sequence alterations of chromatin remodeling genes, *ARID1A* and *ARID1B*, were identified in 8 of 71 tumors (11%) and were associated with early treatment failure and decreased survival. Using tumor-specific structural alterations, we developed an approach to identify rearranged DNA fragments in sera, providing personalized biomarkers for minimal residual disease detection and monitoring. These results highlight dysregulation of chromatin remodeling in pediatric tumorigenesis and provide new approaches for the management of neuroblastoma patients.

Neuroblastomas are pediatric tumors arising from neural crest-derived precursors of the peripheral sympathetic nervous system. As is typical of embryonal tumors, they arise early in childhood with 90% of all cases diagnosed before the age of 5 years. They are the most common extra-cranial solid tumor of childhood and are responsible for up to 15% of childhood cancer-related deaths^{1–3}, with the majority of patients presenting with metastatic disease at the time of diagnosis. Neuroblastomas manifest marked heterogeneity in clinical outcome. The prognosis of children less than 18 months old, even those with metastatic disease, is favorable, and the tumors in children with stage 4S disease frequently regress spontaneously⁴. Unfortunately, children older than 18 months old who are diagnosed with advanced stage disease have a grave prognosis despite multimodal, dose-intensive chemoradiotherapy⁵. Several recurrent genetic alterations have been elucidated, including amplification of the *MYCN* oncogene in ~20% of cases^{6,7}, activating mutations in the *ALK* tyrosine kinase in ~8% of primary tumors^{8–11}, and more recently mutations in *ATRX* in neuroblastomas presenting in older children and adolescents¹². *MYCN* amplification is associated with advanced tumors and poor outcome, *ATRX* mutations define indolent neuroblastoma with eventual progression, while the prognostic value of *ALK* alterations remains to be defined⁷.

RESULTS

Whole-exome and whole-genome next generation sequencing analyses

To comprehensively analyze acquired genetic alterations in neuroblastoma, we used a combination of next generation sequencing approaches in a discovery screen: low-coverage whole-genome sequencing for detection of structural and copy number alterations in 26 cases; exome sequencing for detection of subtle sequence alterations in 16 cases; and high-coverage whole-genome sequencing for detection of both sequence and structural alterations in 6 cases (all of which were also subjected to exome sequencing) (Supplementary Fig. 1, Table 1). In total, 16 cases could be analyzed for subtle mutations such as single base substitutions and small insertions or deletions (indels), while 32 cases (26 with low coverage, 6 with high coverage) could be analyzed for large scale structural changes and copy number alterations. DNA was obtained from low-passage cell lines (n=6) or primary tumors (n=29) and matched normal controls as indicated in Supplementary Table 1. Following library construction and capture on a SureSelect (Agilent) Enrichment System, DNA was sequenced using Illumina GAIIx/HiSeq instruments (Supplementary Note). The average coverage of each base in the targeted regions was 31-fold and 94-fold for the high-coverage whole-genome and exome sequencing approaches, respectively (Supplementary Tables 2 and 3), while the low-coverage whole genomic sequencing achieved an average of 10-fold physical coverage (Supplementary Table 4).

The sequencing data were analyzed using stringent criteria to identify somatic single base substitutions, insertions or deletions (indels), and structural alterations (Online Methods). All single base substitutions and indels were confirmed by an independent sequencing method (Online Methods), and only confirmed mutations are included in the analyses described below. With the exception of one tumor, we found that neuroblastoma tumors had an average of 13 (range, 1 to 52) somatically acquired single base substitution or indel mutations that would be predicted to result in non-silent (NS) changes in coding regions. The NS substitutions were predominantly C:G to A:T transversions (Fig. 1; Supplementary Table 5), representing a mutation spectra different from other pediatric and adult tumors^{13,14,15}. Overall, we detected 368 mutations in 353 genes (Supplementary Table 5). The average number of somatic mutations in neuroblastomas was similar to that reported for neuroblastoma by Molenaar¹⁶ and slightly higher than the number in medulloblastomas, a pediatric tumor analyzed by exome sequencing¹³. This is notably lower than the number of alterations observed in most common adult solid tumors^{14,15}. One tumor-derived cell line, NB07C, had a substantially higher number of somatic mutations (169 NS changes) than the other neuroblastomas analyzed. This case was considered to be an outlier in this study but may identify a unique subset of cases if similar tumors are identified in future validation efforts.

Six samples were analyzed by both exome and high-coverage whole-genome sequencing, permitting independent validation of the somatic alterations as well as a comparison of these approaches for the detection of sequence alterations. Over 91% of the whole-genome and 94% of whole-exome targeted bases were represented by at least 10 reads (Supplementary Tables 2 and 3). A total of 245 somatic alterations in coding regions were detected by either approach with 219 mutations identified by whole-genome sequencing and 240 alterations identified by whole-exome sequencing. Exomic and genomic sequencing detected 98% and 89%, respectively, of the mutations, consistent with similar comparisons made by others¹⁷.

In addition to the single base substitutions and indels, we analyzed copy number changes corresponding to focal amplifications (5-fold copy number gain) or homozygous deletions (less than 20 Mb in size) as these are likely to harbor potential oncogenes and tumor suppressor genes. There was an average of two such focal copy number changes per tumor (range, 0 to 10 per tumor) whose boundaries included at least one protein-encoding gene (Supplementary Table 6); all were amplification events and the majority included either *MYCN* or *ALK* as the putative target gene. One tumor amplicon (in NB1395T) harbored *LIN28B*, which is downstream of *MYCN* and a putative neuroblastoma oncogenic driver^{18,19}. There were also four structural rearrangements per tumor that were within protein-encoding genes (range, 0 to 18 per tumor; Supplementary Tables 4 and 7 and Supplementary Fig. 2). These included deletions, duplications, and inversions within the same chromosome as well as inter-chromosomal translocations. We did not find evidence of chromothripsis in these samples, although this has recently been reported in a subset of high-risk neuroblastoma tumors¹⁶.

Candidate neuroblastoma driver genes and targeted sequencing analyses

The coding exons of all genes that were recurrently altered in the tumors analyzed by next generation sequencing were examined by PCR and Sanger sequencing in 74 additional neuroblastoma cases (Table 2, Supplementary Table 1 and Online Methods). Integration of these data with next generation sequencing data revealed a number of novel genes as well as those previously known to be involved in neuroblastoma. The *ALK* receptor tyrosine kinase gene was found to be mutated in 8 of 90 cases (9%) in our discovery screen (Table 2 and Supplementary Table 5). All eight sequence changes in *ALK* affected two amino acid residues in the tyrosine kinase domain (R1275Q, R1275L and F1174L) that have been reported to lead to constitutive kinase activity^{4,8,11}. An additional 15-fold amplification of

the *ALK* gene was identified in one of 32 cases evaluated for structural changes and copy number alterations (Supplementary Table 6). However, no *ALK* translocations were detected, suggesting that this mechanism of *ALK* activation, typical of large cell lymphomas, non-small cell lung cancers, and inflammatory myofibroblastic tumors, is uncommon in neuroblastoma^{20,21}. Additionally, the *MYCN* oncogene was found to be focally amplified in 15 of the 32 (47%) neuroblastomas, including 5 of the 6 neuroblastoma cell lines, consistent with the previously reported frequency of *MYCN* amplification in high risk tumors and cell lines derived from such tumors⁷ (Table 2 and Supplementary Table 6). Co-amplification of *ODC1*, a *MYCN* target gene important for oncogenicity in neuroblastoma²², was seen in 3 of 15 (20%) *MYCN* amplified tumors (none of which displayed copy number changes of *ALK*). Other alterations in known cancer genes included a glutamine to lysine change at codon 61 in the *HRAS* oncogene, and single missense alterations in the *PTCH1* tumor suppressor and in the EGF receptor family member *ERBB4* (Supplementary Table 5).

In addition to these alterations, a number of mutations in genes not previously known to be involved in neuroblastoma were identified. The most prominent example was the detection of intragenic hemizygous deletions targeting the AT rich interactive domain 1B gene, *ARID1B*, in three of 32 tumors (9%) in the discovery screen (Fig. 2, Table 2, and Supplementary Table 7). The deletions in *ARID1B* were identified by virtue of their aberrantly spaced paired-end sequences and, due to their small size and hemizygous nature, would have been difficult to detect using conventional copy number analyses. These included an 83 kb deletion encompassing exon 6 and a 147 kb deletion encompassing exons 6–9 that were predicted to result in a frameshift and premature truncation of the gene products, and a 621 kb deletion that removed exons 1 and 2, including the protein translation start site (Fig. 2 and Table 2). All these deletions, which were confirmed by PCR amplification and sequencing across the deletion junction, would be expected to abolish functional translation of the key downstream DNA binding (ARID) and topoisomerase-II associated (PAT1) protein domains of *ARID1B*. An additional tumor had an insertion mutation in the homologous *ARID1A* gene that would be predicted to lead to premature termination of the protein.

To investigate the prevalence of these specific alterations identified in the discovery screen, we designed a custom capture approach to selectively sequence and detect point mutations and structural alterations in the genomic regions of *ARID1A*, *ARID1B*, *ALK* and *MYCN* in 40 additional neuroblastoma cases (Supplementary Fig. 1, Prevalence Screen). These analyses yielded an average sequence coverage of 723-fold per targeted base (Supplementary Tables 1 and 8). Through these analyses we were able to identify an intragenic hemizygous deletion, a splice-site mutation and a missense mutation in *ARID1B* in two additional tumors as well as an additional intragenic deletion in a previously analyzed sample (NB05) (Fig. 2, Table 2 and Supplementary Tables 5 and 7). Collectively, *ARID1B* point mutations or intragenic deletions were identified in 5/71 (7%) of neuroblastoma cases (Fig. 2 and Table 2). We further identified hemizygous deletions encompassing the entire coding region of *ARID1B* in the distal region of 6q in 5 additional cases (Supplementary Table 6). Furthermore, point mutations of *ARID1A* were identified in three additional cases, two of which led to biallelic inactivation through mutation predicted to result in premature termination of the protein and deletion of the alternative allele at 1p36 (Fig. 2 and Table 2, Supplementary Table 5). All of these alterations were confirmed by Sanger sequencing. Not surprisingly, we identified additional *ALK* missense changes and *MYCN* amplifications, resulting in somatic alterations of *ALK* in 18/130 (14%) and of *MYCN* in 43/71 (61%) of total cases (Table 2, Supplementary Tables 5 and 6).

ARID1B is a member of the SWI/SNF transcriptional complex that is thought to regulate chromatin structure²³. Mutations recently identified in *ARID1B* suggest that it may serve as a potential tumorigenic driver in a small fraction of hepatocellular²⁴, breast²⁵, ovarian²⁶, and medulloblastoma^{27,28} tumors. Through our integrated genomic analyses, our findings of five independent structural alterations and two sequence changes, the majority of which would result in a truncated protein, strongly support this gene as a contributor to neuroblastoma oncogenesis (passenger probability $P < 0.001$). Interestingly, we found sequence alterations in other genes involved in chromatin regulation in neuroblastoma. These included two frameshift, one nonsense and one missense mutation in *ARID1A*, another SWI/SNF complex member, nonsense mutations in the histone acetyl transferase (HAC) genes *EP300* and *CREBBP*, and missense mutations in the SWI2/SNF2 family member *TTF2* gene, the histone demethylase gene *KDM5A*, and the chromatin remodeling zinc finger gene *IKZF1*. Genes involved in chromatin structure or remodeling have been reported to be implicated in human cancers. These include a high frequency of alterations of *ARID1A* in ovarian clear cell carcinomas²⁶, *SMARCB1* in malignant rhabdoid tumors²⁹, alterations of *PBRM1* in renal cell carcinomas³⁰, alterations of *EP300* and *CREBBP* in transitional cell carcinomas of the bladder³¹ and B cell lymphomas³², alterations of *DAXX* and *ATRX* in pancreatic endocrine tumors³³, and inactivation of histone methyltransferases *MLL2* and *MLL3* in medulloblastomas¹³, among others^{34–36}. Of note, *ATRX* has recently been shown to be mutated in neuroblastoma tumors from adolescents and young adults (12 years old)¹² but would not have been expected to be altered in a significant fraction of the patients evaluated in our study (median age of diagnosis <2 years old, range <1 to 6 years old).

Personalized genomic biomarkers for neuroblastoma patients

Although the number of sequence alterations in neuroblastomas was low compared to adult tumors, the frequency of recurrent structural rearrangements in neuroblastomas was relatively high. Every tumor had at least one rearrangement (range, 1 to 66) and all cases that had recurrent copy number changes of the *MYCN*, *ARID1B*, or *ALK* genes also had rearrangements at these loci. Such rearrangements are not present in normal cells and could therefore be useful as biomarkers of neuroblastoma. Given the poor treatment outcomes of many neuroblastoma patients, the availability of non-invasive biomarkers to detect minimal residual disease after surgery and to measure molecular response to chemotherapy would be useful for clinical management of neuroblastoma patients.

To demonstrate the feasibility of this approach, we developed personalized biomarkers based on the rearrangements present in the cancers analyzed³⁷. This was performed through analysis of either whole-genome sequencing or capture and sequencing of the *MYCN* locus to identify structural alterations associated with novel rearrangement junctions not present in the germline (Online Methods). We have previously shown that tumor-specific rearrangements have the potential to serve as highly sensitive biomarkers for tumor detection and monitoring³⁷, and would therefore be expected to have fundamental advantages over measurement of wild-type sequences, including wild-type *MYCN* levels³⁸, in neuroblastoma patients. Notably, both *MYCN* amplified and non-amplified tumors had identifiable somatic rearrangement biomarkers, and in three cases in which serum was available at the time of diagnosis, we were able to detect and quantify such specific tumor rearrangements in the patients' serum (Table 3, Supplementary Table 9). Interestingly, quantitative analyses showed that there was much more tumor DNA freely floating in the serum than in circulating cells, suggesting that the cell free compartment of blood may represent a more sensitive source for detection of tumor burden (Table 3).

We developed personalized rearrangement biomarkers to monitor circulating tumor DNA (ctDNA) in serial plasma samples from four additional cases of neuroblastoma obtained during a post-consolidation minimal residual disease (MRD) immunotherapy trial³⁹

(Supplementary Fig. 3). In two cases, NB2885T and NB2870T, the ctDNA was detected at the end of standard high risk neuroblastoma therapy and, despite MRD immunotherapy, went on to relapse and eventually die of disease. The prolonged reduction in ctDNA in NB2885T during immunotherapy may be an indication of therapeutic response whereas the marked increase in ctDNA in NB2870T correlated with clinical relapse during the trial period. In cases NB6321T and NB2464T, no ctDNA was detectable and these patients were alive at the last follow-up over one and four years later, respectively. These data demonstrate that ctDNA may be a useful surrogate for the level of clinical disease, and that the presence of ctDNA may be a highly sensitive and specific predictor of minimal residual disease and subsequent relapse⁴⁰.

ARID1 alterations and clinical correlates

These genome-wide sequence analyses suggest that neuroblastoma tumors are driven by a relatively small number of somatically acquired alterations and that genes involved in chromatin remodeling, including *ARID1B* and *ARID1A*, were enriched for alterations. ARID1 family genes are integral components of the SWI/SNF neural progenitors-specific chromatin remodeling BAF complex that is essential for the self-renewal of multipotent neural stem cells⁴¹. Tumor-specific deletions encompassing *ARID1B* have been reported in CNS tumors⁴² and multiple members of this complex have been identified as tumor suppressor genes^{26,41}. We found that high expression of members unique to the neural-progenitor BAF complex correlates with a high-risk neuroblastoma phenotype while high expression of those specific to the neuron specific BAF complex, or downstream neuritogenesis target genes, correlates with lower risk neuroblastoma (Supplementary Fig. 4). These data support a model whereby disrupted BAF complex signaling may preserve an undifferentiated progenitor state.

The model above would suggest that alterations in ARID1 may correlate with a more aggressive neuroblastoma phenotype. All but one of the patients with alterations in *ARID1A* or *ARID1B* died of progressive disease, including a child with low-risk neuroblastoma (a group with a survival probability of >98%). ARID1 alterations were associated with inferior overall survival of 386 days compared to 1689 days for patients without such alterations (hazard ratio, HR 4.49; 95% confidence interval, CI 1.24–16.33; $P=0.0226$, log-rank test; Fig. 3 and Supplementary Table 10). An analysis that also included hemizygous deletions of the entire coding region of *ARID1B* further increased the significance of the survival difference between patients with mutant and wildtype *ARID1B/A* (hazard ratio, HR 6.41; 95% confidence interval, CI 1.93–21.25; $P=0.0024$, log-rank test). The median survival of patients with ARID1 alterations was lower than that of any other genetic alterations assessed, including *MYCN* amplification (median survival 726 days) providing a potential marker for early therapy failure and disease progression.

DISCUSSION

This study underlies the importance of integrated genomic analyses, including detection of sequence alterations, copy number changes, and rearrangements that can now be performed using massively parallel sequencing approaches to identify subtle genomic changes. Despite the comprehensive efforts of this study, some alterations may not have been detected. First, a small fraction of the exome was not analyzed, either due to low sequence coverage in the whole-genome analyses or inadequate capture in the exome analyses. Second, it is possible that point mutations in non-protein-coding regions of the genome may be involved in neuroblastoma. Such data were obtained for six neuroblastoma cases and did not identify any clear clustering of alterations; analysis of additional neuroblastoma cases could be useful to further interpret these non-coding changes. Third, germline neuroblastoma susceptibility variants have been identified^{43,44} and additional such variants yet to be

discovered may be present in our neuroblastoma cases. Fourth, it is possible that epigenetic alterations contribute to the initiation or progression of neuroblastomas. This possibility is intriguing given the new data on *ARID1B* and *ARID1A* in this tumor type. Finally, although rearrangements and copy number changes were detected in a genome-wide fashion, many of these occurred in non-coding regions and their functional roles remain to be elucidated.

Our data add to the growing knowledge of the genomic landscapes of human cancers. They are consistent with the idea that pediatric tumors do not require as many genetic alterations as typical adult cancers^{13,45}. Although few alterations were identified in known therapeutically-targetable oncogenes such *ALK*, there are many other alterations, both subtle and large, that are found in these cancers and many of these affect chromatin-modifying genes. These data highlight the important connection between genetic alterations in the cancer genome and epigenetic pathways, and provide new avenues for research and disease management in neuroblastoma patients.

ONLINE METHODS

Samples Obtained for Sequencing Analyses

Neuroblastoma tumor DNA (from cell lines and primary tumors), matched germline DNA (from peripheral blood or lymphoblastoid cell line) and patient serum or plasma were obtained from the Children's Oncology Group (COG) cell line repository and the COG Neuroblastoma biobank following committee approval (study #COG NB 2008-02). Informed consent for research use was obtained from all patients and/or parents at the enrolling COG member institution prior to tissue banking or cell line generation, and study approval was obtained from The Children's Hospital of Philadelphia Institutional Review Board. All samples were STR genotyped to confirm identity. Primary tumor samples were selected from patients with COG high-risk disease, and specimens verified to have >75% viable tumor cell content by histopathology assessment. Serial plasma samples for MRD assays were obtained from patients enrolled on the COG ANBL0032 immunotherapy study.

Massively Parallel Paired-End Sequencing and Somatic Mutation Identification

Genomic DNA libraries were prepared and captured following Illumina's (Illumina, San Diego, CA) suggested protocol with the modifications described in the Supplementary Note, or by Personal Genome Diagnostics (Baltimore, MD). DNA libraries were sequenced with the Illumina GAIIX/HiSeq Genome Analyzer, yielding 100 or 200 base pairs of sequence from the final library fragments for high coverage exome/low coverage genome and high coverage genome analyses respectively. Sequencing reads were analyzed and aligned to human genome hg18 with the Eland algorithm in CASAVA 1.7 software (Illumina). Reads were mapped using the default seed-and-extend algorithm, which allowed a maximum of 2 mismatched bases in the first 32bp of sequence. Identification of somatic alterations was performed as previously described^{46–49} utilizing a next-generation sequencing analysis pipeline that enriched for tumor-specific single nucleotide alterations and small insertions/deletions. Briefly, for each position with a mismatch (as compared to the hg18 reference sequence using the Eland algorithm) the read coverage of the mismatch and wild-type sequence at that base was calculated. A candidate mismatched base was identified as a mutation only when (i) two or more distinct paired-tags contained the mismatched base; (ii) the number of distinct paired-tags containing a particular mismatched base was at least 7.5% of the total distinct tags; and (iii) the mismatched base was not present in >0.5% of the tags in the matched normal sample. Candidate somatic point mutations identified by next generation sequencing approaches were confirmed by an independent sequencing method (either a different next-generation sequencing approach or polymerase chain reaction (PCR) followed by Sanger sequencing, Supplementary Table 5).

Evaluation of Genes in Additional Tumors and Matched Normal Controls

For 12 selected genes that were somatically altered, the coding region was sequenced in a validation set composed of an independent series of 74 additional neuroblastomas and matched controls. These genes included *ALK*, *ANKRD34B*, *ARID1B*, *ARID1A*, *FAR1*, *PRSS16*, *PRSS23*, *RASGRP3*, *TLL6*, *VANGL1*, *VCAN* and *ZHX2*. PCR amplification and Sanger sequencing analyses were performed following protocols described previously¹⁵.

Identification of Somatic Copy Number Alterations

Single tags passing filter were grouped by genomic position in nonoverlapping 3-kb bins. A tag density ratio was calculated for each bin by dividing the number of tags observed in the bin by the average number of tags expected to be in each bin (on the basis of the total number of tags obtained for chromosomes 1 to 22 for each library divided by 849,434 total bins). The tag density ratio thereby allowed a normalized comparison between libraries containing different numbers of total tags. A control group of libraries made from the six matched normal high coverage whole-genome samples from Supplementary Table 1 and six additional normal samples [Co84N, Co108N, B5N, B7N³⁷ and CEPH (Centre d'Etude du Polymorphisme Humain) samples NA07357 and NA18507] was used to define areas of germline copy number variation or that contained a large fraction of repeated or low-complexity sequences. Any bin where at least two of the normal libraries had a tag density ratio of <0.25 or >1.75 was removed from further analysis.

For all samples analyzed with low coverage whole-genome sequencing (Supplementary Table 4), amplifications were identified as three or more bins with tag ratios of >2 , separated by no more than ten intervening bins with a tag ratio <2 . For all amplifications, at least one bin had a tag ratio of ≥ 5 . For samples with high coverage whole-genome sequencing (Supplementary Table 3), homozygous deletions were identified as three or more bins with tag ratios of <0.25 , separated by no more than ten intervening bins with a tag ratio >0.25 . Single-copy gains and losses were identified through visual inspection of tag density data for each sample.

For all samples analyzed with targeted capture sequencing, the tag ratio for each gene was calculated as the average read coverage for the gene, divided by the average read coverage of the *ALK*, *ARID1A* and *ARID1B* genes (*MYCN* was not used as it is frequently amplified). These values were normalized to the average coverage for each gene in a normal sample. Amplifications and hemizygous deletions were identified if the tag ratio for a gene was ≥ 5.0 or < 0.65 , respectively. Hemizygous deletions were confirmed through LOH analyses of SNPs in the genomic region of each gene.

Six samples with high coverage whole-genome sequencing were analyzed for amplifications at the *MYCN* locus. The boundary coordinates for these amplifications were compared and a one megabase (hg18 chr2:15.5Mb–16.5Mb) region was identified that contained at least one amplification boundary region from each sample.

Identification of Somatic Rearrangements

Somatic rearrangements were identified by querying aberrantly mapping reads from one flow cell of an Illumina GAIIx run (100bp PE) or up to two lanes of an Illumina HiSeq Genome Analyzer run (50bp PE) to achieve a physical coverage of $>8X$. The discordantly mapping pairs were grouped into 1kb bins when at least 2 distinct tag pairs (with distinct start sites) spanned the same two 1kb bins (known bins which contained aberrantly mapping tags were removed as described above³⁷, as well as 1kb bins involved in known germline structural alterations⁵⁰).

To identify all high-confidence genomic rearrangements, candidate rearrangements were filtered using the above described criteria and were required to have at least one tag sequenced across the rearrangement breakpoint. Breakpoints were determined using BLAT alignment to the human genome sequence (hg18)⁵¹. In order to ensure that no recurrent rearrangements in coding genes were missed, genes which harbored rearrangements were evaluated for all candidate rearrangements without the requirement that the breakpoint be present in a sequenced tag and any recurrent gene rearrangement was further analyzed. Candidate rearrangements were confirmed as somatic when a 10 uL PCR based reaction (containing 5.9 uL H₂O, 1 uL 10X PCR buffer, 1 uL 10mM dNTPs, 0.6 uL DMSO, 0.4 uL 25uM primers, 0.1 uL Platinum Taq and 1 uL DNA, 3 ng/uL) resulted in the amplification of a product of the expected size in the tumor but not in the matched normal on a 1% ethidium bromide stained agarose gel. Utilizing this stringent pipeline, of the 26 candidate genomic rearrangements tested, 25 were confirmed as somatic (96%) as well as 15 of the 16 candidate rearrangements tested that were identified by the *NMYC* capture sequencing method (94%). In all three cases of *ARID1B* somatic rearrangement, the PCR product was Sanger sequenced to identify the breakpoint to the base-pair resolution. For biomarker analyses, rearrangements were identified with the initial-above described method, with a subsequent PCR product sequenced and aligned using BLAT to hg18⁵¹ in order to design primers to amplify a PCR product in the serum, plasma or peripheral blood between 70 and 120 bp.

Quantification of Tumor Burden in Serum and Peripheral Blood

Circulating tumor DNA was amplified using 2x Phusion Flash PCR Master Mix and patient specific primers (at a final concentration of 0.5uM each) in DNA isolated from serum or plasma and DNA isolated from peripheral blood cells. Subsequently, the level of tumor DNA was quantified after amplification by digital PCR on SYBR green I stained 10% TBE gels³⁷.

Gene Expression Analyses

For gene expression profiling by Affymetrix U95Av2 microarrays, the expression measures for each probe set was extracted and normalized using robust multi-array average protocols from raw CEL files as described previously. Basic linear correlation and regression was used to define r , r^2 and two-tailed p value to assess correlation among gene expression values.

Statistical Analyses for Clinical and Genetic Data

Curves for overall survival (calculated as the time from diagnosis) were constructed using the Kaplan-Meier method and compared between groups using the log-rank test for descriptive purposes. Cox proportional hazards models were used to test for the effect of clinical and genetic parameters on survival. Passenger probabilities were calculated using the binomial test adjusted for gene sizes and corrected for multiple comparisons⁵².

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the families and children with neuroblastoma who contributed to this work. We thank John Maris for valuable input to this work, J. Ptak, N. Silliman, L. Dobbyn, M. Whalen, J. Schaefer, and T. Mosbrugger for technical assistance with sequencing analyses, Lisa Kann and Sam Angiuoli of Personal Genome Diagnostics for targeted sequence analyses, the Children's Oncology Group (COG), Wendy B. London and the COG Statistics and Data Center, Julie Gastier-Foster and the Neuroblastoma Reference Laboratory, Nilsa Ramirez and the

Biopathology Center, Cindy Winter and the CHOP Nucleic Acids Bank, and Tito Woodburn and the COG Cell Line Repository. This work was generously supported by the St. Baldrick's Foundation for childhood cancer research, the Virginia and D. K. Ludwig Fund for Cancer Research, Swim Across America, AACR Stand Up To Cancer-Dream Team Translational Cancer Research Grant, and NIH grant CA121113.

References

1. Ries, LAG.; SM; Gurney, JG.; Linet, M.; Tamra, T.; Young, JL.; Bunin, GR. NIH Publication. Vol. 99-4649. National Cancer Institute, SEER Program; Bethesda, MD: 1999. Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975–1995.
2. Maris JM. Recent advances in neuroblastoma. *N Engl J Med*. 2010; 362:2202–11. [PubMed: 20558371]
3. Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. *Lancet*. 2007; 369:2106–20. [PubMed: 17586306]
4. Capasso M, Diskin SJ. Genetics and genomics of neuroblastoma. *Cancer Treat Res*. 2010; 155:65–84. [PubMed: 20517688]
5. Mueller S, Matthay KK. Neuroblastoma: biology and staging. *Curr Oncol Rep*. 2009; 11:431–8. [PubMed: 19840520]
6. Schwab M, et al. Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature*. 1983; 305:245–8. [PubMed: 6888561]
7. Brodeur GM, Seeger RC. Gene amplification in human neuroblastomas: basic mechanisms and clinical implications. *Cancer Genet Cytogenet*. 1986; 19:101–11. [PubMed: 3940169]
8. Chen Y, et al. Oncogenic mutations of ALK kinase in neuroblastoma. *Nature*. 2008; 455:971–4. [PubMed: 18923524]
9. George RE, et al. Activating mutations in ALK provide a therapeutic target in neuroblastoma. *Nature*. 2008; 455:975–8. [PubMed: 18923525]
10. Janoueix-Lerosey I, et al. Somatic and germline activating mutations of the ALK kinase receptor in neuroblastoma. *Nature*. 2008; 455:967–70. [PubMed: 18923523]
11. Mosse YP, et al. Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature*. 2008; 455:930–5. [PubMed: 18724359]
12. Cheung NK, et al. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. *JAMA*. 2012; 307:1062–71. [PubMed: 22416102]
13. Parsons DW, et al. The genetic landscape of the childhood cancer medulloblastoma. *Science*. 2011; 331:435–9. [PubMed: 21163964]
14. Jones S, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. 2008; 321:1801–6. [PubMed: 18772397]
15. Sjoblom T, et al. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006; 314:268–74. [PubMed: 16959974]
16. Molenaar JJ, et al. Sequencing of neuroblastoma identifies chromothripsis and defects in neurogenesis genes. *Nature*. 2012; 483:589–93. [PubMed: 22367537]
17. Clark MJ, et al. Performance comparison of exome DNA sequencing technologies. *Nat Biotechnol*. 2011
18. Viswanathan SR, et al. Lin28 promotes transformation and is associated with advanced human malignancies. *Nat Genet*. 2009; 41:843–8. [PubMed: 19483683]
19. Cotterman R, Knoepfler PS. N-Myc regulates expression of pluripotency genes in neuroblastoma including *lif*, *klf2*, *klf4*, and *lin28b*. *PLoS One*. 2009; 4:e5799. [PubMed: 19495417]
20. Rikova K, et al. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell*. 2007; 131:1190–203. [PubMed: 18083107]
21. Soda M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature*. 2007; 448:561–6. [PubMed: 17625570]
22. Hogarty MD, et al. ODC1 is a critical determinant of MYCN oncogenesis and a therapeutic target in neuroblastoma. *Cancer Res*. 2008; 68:9735–45. [PubMed: 19047152]

23. Wang X, et al. Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochem J.* 2004; 383:319–25. [PubMed: 15170388]
24. Fujimoto A, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet.* 2012
25. Stephens PJ, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature.* 2012; 486:400–4. [PubMed: 22722201]
26. Jones S, et al. Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science.* 2010; 330:228–31. [PubMed: 20826764]
27. Jones DT, et al. Dissecting the genomic complexity underlying medulloblastoma. *Nature.* 2012; 488:100–5. [PubMed: 22832583]
28. Pugh TJ, et al. Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature.* 2012; 488:106–10. [PubMed: 22820256]
29. Versteeg I, et al. Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature.* 1998; 394:203–6. [PubMed: 9671307]
30. Varela I, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature.* 2011; 469:539–42. [PubMed: 21248752]
31. Gui Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet.* 2011; 43:875–8. [PubMed: 21822268]
32. Pasqualucci L, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature.* 2011; 471:189–95. [PubMed: 21390126]
33. Jiao Y, et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science.* 2011; 331:1199–203. [PubMed: 21252315]
34. Schwartzentruber J, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature.* 2012; 482:226–31. [PubMed: 22286061]
35. Wu G, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat Genet.* 2012; 44:251–3. [PubMed: 22286216]
36. Robinson G, et al. Novel mutations target distinct subgroups of medulloblastoma. *Nature.* 2012; 488:43–8. [PubMed: 22722829]
37. Leary RJ, et al. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med.* 2010; 2:20ra14.
38. Combaret V, et al. Circulating MYCN DNA as a tumor-specific marker in neuroblastoma patients. *Cancer Res.* 2002; 62:3646–8. [PubMed: 12097268]
39. Yu AL, et al. Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. *N Engl J Med.* 2010; 363:1324–34. [PubMed: 20879881]
40. Diehl F, et al. Circulating mutant DNA to assess tumor dynamics. *Nat Med.* 2008; 14:985–90. [PubMed: 18670422]
41. Ho L, Crabtree GR. Chromatin remodelling during development. *Nature.* 2010; 463:474–84. [PubMed: 20110991]
42. Ichimura K, et al. Small regions of overlapping deletions on 6q26 in human astrocytic tumours identified using chromosome 6 tile path array-CGH. *Oncogene.* 2006; 25:1261–71. [PubMed: 16205629]
43. Capasso M, et al. Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. *Nat Genet.* 2009; 41:718–23. [PubMed: 19412175]
44. Wang K, et al. Integrative genomics identifies LMO1 as a neuroblastoma oncogene. *Nature.* 2011; 469:216–20. [PubMed: 21124317]
45. Knudson AG Jr, Meadows AT, Nichols WW, Hill R. Chromosomal deletion and retinoblastoma. *N Engl J Med.* 1976; 295:1120–3. [PubMed: 980006]
46. Wu J, et al. Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proc Natl Acad Sci U S A.* 2011; 108:21188–93. [PubMed: 22158988]
47. Bettegowda C, et al. Mutations in CIC and FUBP1 Contribute to Human Oligodendroglioma. *Science.* 2011; 333:1453–1455. [PubMed: 21817013]

48. Wu J, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. *Sci Transl Med*. 2011; 3:92ra66.
49. Agrawal N, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science*. 2011; 333:1154–7. [PubMed: 21798897]
50. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–73. [PubMed: 20981092]
51. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res*. 2002; 12:656–64. [PubMed: 11932250]
52. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*. 1995; 57:289–300.

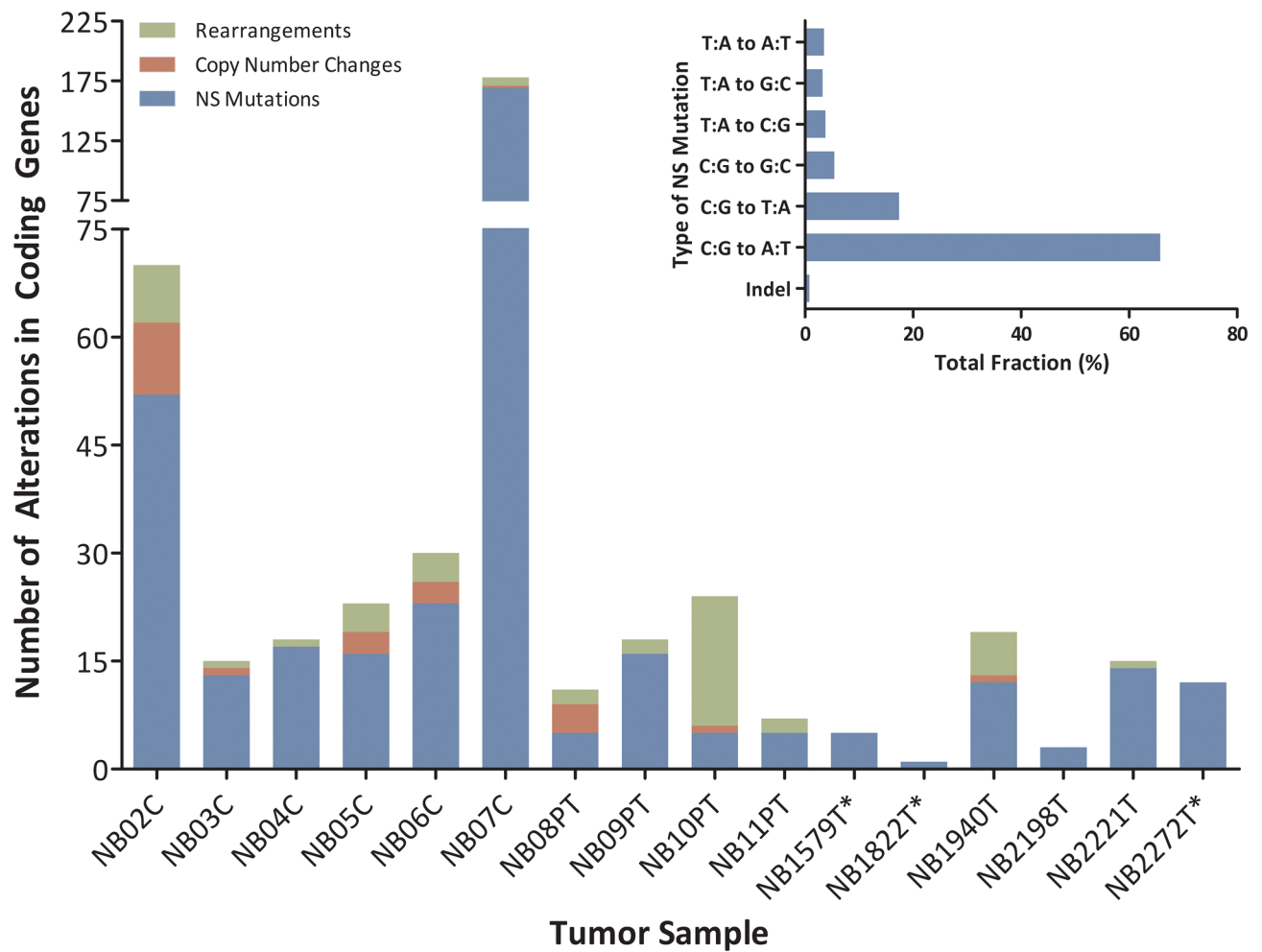


Figure 1. Number and type of somatic alterations detected in each neuroblastoma case
 The vertical axis includes non-synonymous single base substitutions, insertions, deletions, and splice site changes (NS Mutations), homozygous deletions and amplifications affecting protein encoding genes, and rearrangements with at least one breakpoint within the coding region of a gene. The inset shows the mutation spectra of somatic non-silent single nucleotide mutations in 16 cases of neuroblastoma. Data on rearrangements and copy number changes were not available for starred samples.

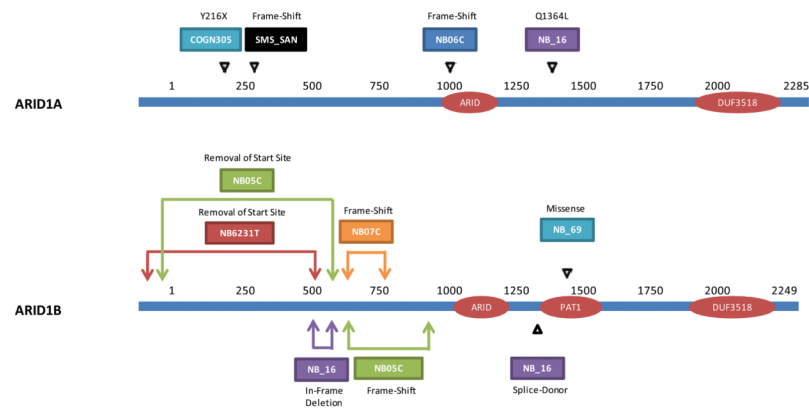


Figure 2. Genomic alterations in *ARID1A* and *ARID1B*

The schematic represents the ARID1B and ARID1A proteins with the predicted effects of observed intragenic deletions and point mutations.

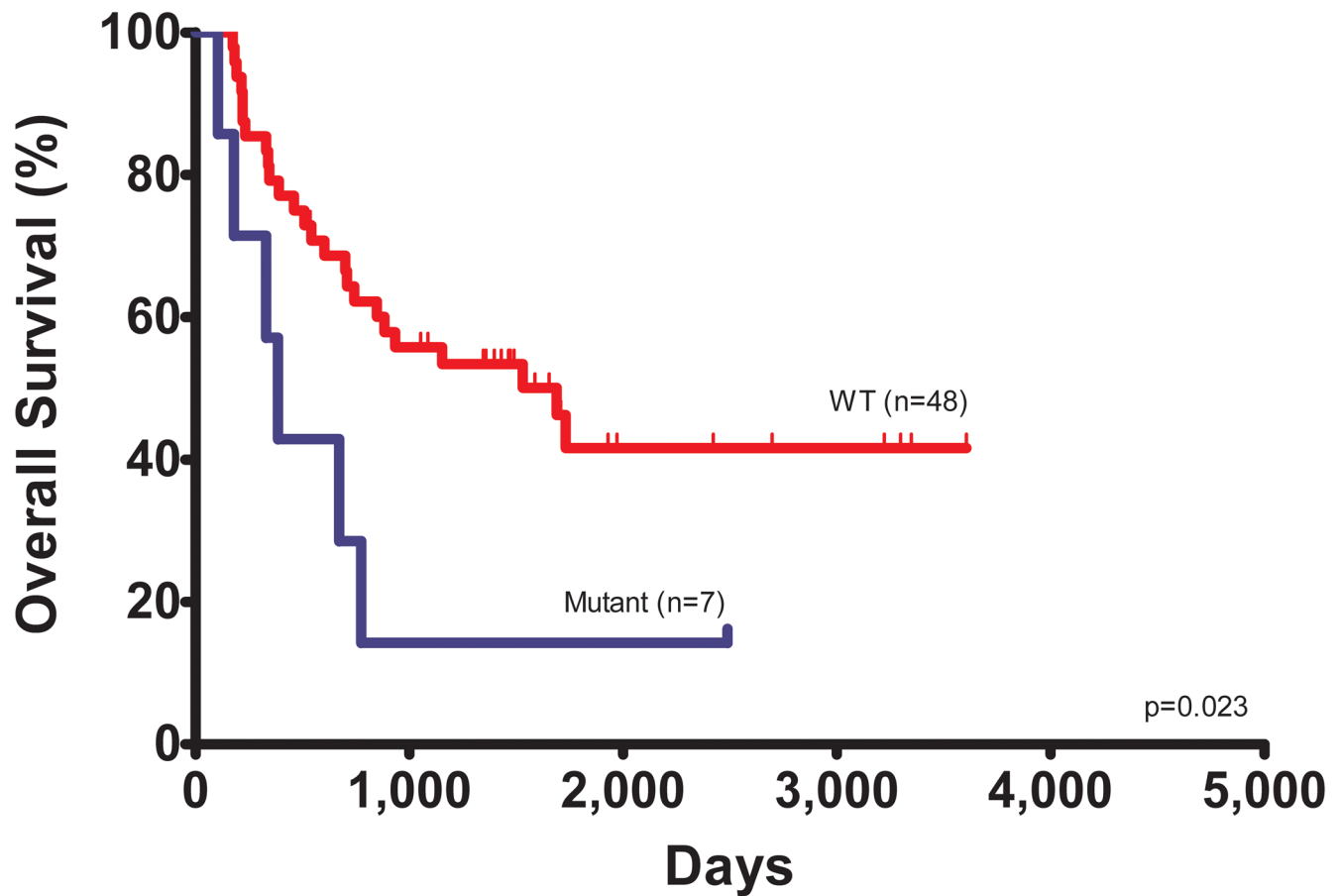


Figure 3. Overall survival according to ARID1B status

The hazard ratio for death among patients with wildtype *ARID1B/A* (n=48), as compared to those with mutant *ARID1B/A* (n = 7) was 4.49 (95% confidence interval, CI 1.24–16.33; $P=0.0226$, log-rank test). The median survival was 1689 days for patients with wildtype *ARID1B/A* compared to 386 days for patients with mutated *ARID1B/A*. An analysis that also included hemizygous deletions of the entire coding region of *ARID1B* further increased the significance of the survival difference between patients with mutant and wildtype *ARID1B/A* (hazard ratio, HR 6.41; 95% confidence interval, CI 1.93–21.25; $P=0.0024$, log-rank test).

Table 1

Summary of next generation sequencing analyses in neuroblastoma *

Sequencing Analysis	Samples Analyzed	Coverage (fold)	Average High Quality Mapped Bases per Sample	Type of Alteration Detected
Exome	16 tumors and 16 matching normals	94	3,781,568,777	Point Mutations
High-Coverage Whole-Genome	6 tumors and 6 matching normals	31	118,719,178,942	Point Mutations, Copy Number, Rearrangements
Low-Coverage Whole-Genome	26 tumors	10	14,691,665,206	Copy Number, Rearrangements
Genomic regions containing <i>ALK</i> , <i>ARID1A</i> , <i>ARID1B</i> and <i>MYCN</i>	40 tumors	723	906,665,738	Point Mutations, Copy Number, Rearrangements
Total Distinct Tumors	74 tumors			Point Mutations (55 tumors), Copy Number (71 tumors), Rearrangements (71 tumors)

* For whole-exome, high-coverage whole-genome, and targeted capture analyses, sequence coverage is indicated while for low-coverage whole-genome analyses, physical coverage is indicated.

Table 2

Summary of recurrent genomic alterations observed in neuroblastoma*

Gene	Accession	Cases Affected (%)	Sample(s)	Type of Somatic Alteration	Reference Genome Coordinates (hg18)	Predicted Transcript Effect	Predicted Protein Alteration
<i>MYCN</i>	CCDS1687.1	61% (43/71)	43 cases	Focal Amplification	chr2:15,998,134–16,004,580	Amplification	Amplification
<i>ALK</i>	CCDS33172.1	14% (18/130)	18 cases	Point Mutation/Focal Amplification	chr2:29,269,144–29,997,981	932, 3271, 3521, 3522, 3824	311, 1091, 1174, 1275
<i>ARID1B</i>	CCDS55072.1	7% (5/71)	NB05C	Hemizygous Deletion	chr6:157,376,737–157,523,337 (146,601 bp)	Exon 6, 7, 8 and 9 Deletion (842 bp)	Frame-Shift
			NB05C	Hemizygous Deletion	chr6:157,328,203–157,122,833 (205,371 bp)	Exon 1, 2, 3, 4 and 5 Deletion (2,037 bp)	Removal of Start Site
			NB07C	Hemizygous Deletion	chr6:157,388,358–157,471,373 (83,016 bp)	Exon 6 Deletion (244 bp)	Frame-Shift
			NB6231T	Hemizygous Deletion	chr6: 156,572,360–157,193,843 (621,484 bp)	Exon 1 and 2 Deletion (1,737 bp)	Removal of Start Site
			NB_16	Hemizygous Deletion	chr6:157,159,346–157,207,071 (47,726 bp)	Exon 2 Deletion (195 bp)	In-Frame Deletion
	CCDS5251.1		NB_16	Point Mutation	chr6:157559145A>G	IVS16+4	Splice-Donor
			NB_69	Point Mutation	chr6:157563781C>T	4307C>T	S1436L
<i>ARID1A</i>	CCDS285.1	6% (4/71)	NB06C	Point Mutation/LOH	chr1:26970227msG	3229insG	Frame-Shift
			SMS_SAN	Point Mutation/LOH	chr1:26896234delGCCCTCCCTCCT	Frame-Shift	Frame-Shift
			NB_16	Point Mutation	chr1:26972966A>T	4091A>T	Q1364L
			COGN305	Point Mutation/LOH	chr1:26896129C>A	648C>A	Y216X
<i>VANGL2</i>	CCDS883.1	2% (2/90)	NB04C, NB02C	Point Mutation	chr1:16026617G>T, chr1:16008285G>A	922G>T, 685G>A	G308W, A229T
<i>ZHX3</i>	CCDS6336.1	2% (2/90)	NB06C, NB10PT	Point Mutation	chr8:124035104G>A, chr8:124034190G>A	2173G>A, 1259G>A	D725N, R420H

* Indicated genes were altered in two or more tumors (not including NB07C) by next generation and Sanger sequencing analyses of neuroblastoma cases (passenger probabilities < 0.001, Online Methods). For *MYCN* and *ALK*, the entire genomic locus is listed, with the affected codons and amino acids indicated. Coordinates refer to human reference genome hg18 release (NCBI 36.1, March 2006).

Biomarker Analyses in Neuroblastomas*

Table 3

Tumor Sample	Time Points Analyzed	Sequencing Analysis	Distinct Paired Tags Analyzed	Physical Coverage	Somatic Rearrangements	Mutant Template Molecules in Serum or Plasma (per mL)	Mutant Template Molecules in Circulating Tumor Cells (per mL)	Post-MRD Therapy Outcome
NB02C	At Diagnosis (1)	High Coverage Whole-Genome	154,389,649	15	16	136	40	Not Enrolled
NB04C	At Diagnosis (1)	High Coverage Whole-Genome	155,886,351	16	1	48,700	2,020	Not Enrolled
NB03C	At Diagnosis (1)	<i>MYCN</i> Locus Capture	4,048,315	911	2	185,000	30,200	Not Enrolled
NB2885T	MRD Therapy (7)	Low Coverage Whole-Genome	131,086,400	10	7	< 1.0 – 16.6	ND	Died of Disease
NB2870T	MRD Therapy (2)	Low Coverage Whole-Genome	62,151,315	5	2	811–8,450	ND	Died of Disease
NB2464T	MRD Therapy (7)	Low Coverage Whole-Genome	61,486,874	5	1	< 0.7	ND	Alive at Follow-up
NB6321T	MRD Therapy (3)	Low Coverage Whole-Genome	134,781,854	10	14	< 0.7	ND	Alive at Follow-up

* Distinct paired tags analyzed is based on unique start sites for each paired-end read; Physical coverage is the number of paired reads expected to span any location in a haploid genome; chr2:15.5Mb–16.5Mb was considered for the *MYCN* locus capture sequencing analysis. Where multiple time points were available (as indicated by the number of samples analyzed in parentheses), the range of circulating tumor DNA is shown. ND indicates the sample was not available; MRD, minimal residual disease.